DOCUMENT RESUME

ED 174 102 HE 011 443

AUTHOR Lind, Douglas A.
TITLE Bayesian Decision Theory in Enrollment Forecasting. AIR Forum 1979 Paper.
PUB DATE May 79
NOTE 15p.; Paper presented at the Annual Forum of the Association for Institutional Research (19th, San Diego, California, May 13-17, 1979)

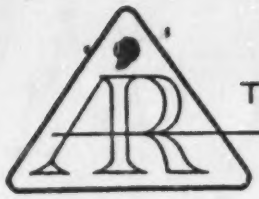
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Enrollment Projections; Higher Education; *Institutional Research; *Mathematical Models; *Prediction; *Probability Theory; Statistical Analysis
IDENTIFIERS *AIR Forum 1979; University of Toledo OH

ABSTRACT

The use of subjective probability as a theoretical model for enrollment forecasting is proposed, and the results of an application of subjective probability to enrollment forecasting at the University of Toledo are reported. Subjective probability can be used as an enrollment forecasting technique for both headcount and full-time equivalent using the following steps: (1) assume the various subpopulations are normally distributed; (2) determine the mean and standard deviations of these subpopulations; (3) sum the means and variances of the subpopulations to obtain the mean and variance of the university forecast; and (4) perform a linear transformation on the headcount distribution to obtain the full-time equivalent distribution. Major advantages of this method are: it allows the study of homogeneous subpopulations; it allows comparisons with other methods, since the variability is known; and it allows probability statements regarding various headcount and full-time equivalent outcomes. Application of the subjective probability method to enrollment forecasting at the University of Toledo resulted in a difference of 0.4 percent between the actual and the forecasted headcount enrollment and a difference of 0.3 percent between actual and forecasted full-time equivalent enrollment. In comparison with the conventional forecasting techniques also employed in the test study, the forecasting errors using the subjective probability were the smallest. (SC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174102



THE ASSOCIATION FOR INSTITUTIONAL RESEARCH

This paper was presented at the Nineteenth Annual Forum of the Association for Institutional Research held at the Town and Country Inn in San Diego, California, May 13-17, 1979. This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC collection of Forum papers.

Mary Corcoran
University of Minnesota
(Editor, AIR Forum Publications)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

AIR

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NE 011 443

Bayesian Decision Theory
in
Enrollment Forecasting

Douglas A. Lind
Assistant Professor
The University of Toledo
2801 W. Bancroft
Toledo, Ohio 43606
(419) 537-2346

ABSTRACT

The purpose of this paper is to suggest the use of subjective probability in enrollment forecasting and to report on the results of a trial at The University of Toledo. The mean and standard deviation for homogenious subpopulations are subjectively estimated and then totaled over the university. Through a linear transformation the headcount means and standard deviations are converted to an FTE basis. The major advantages to the use of subjective probability is that the variance highlights the most uncertain portions of the projection and that probability statements can be made regarding various possible outcomes.

Introduction

During the last ten years there has been an increasing diversity among students in postsecondary education, particualry in the ages at which they enroll, the types of studies they select, and the number who complete and obtain degrees. This increase has been accompanied by a program diversification that has added to the statistical complexity of postsecondary education and, thereby, signaled a caution to those attempting to predict the future enrollment patterns within higher education.

There has been a substantial expansion in higher education, particularly among the public institution, since 1960. However, an enrollment decline has been predicted for the mid-1980's. Educators are concerned about the future, given this prediction. In an effort to maintain enrollment, many universities have recruited nontraditional students (students outside the eighteen to twenty-four year age group). Therefore, the basic assumptions undergirding prior forecasting methodologies may not be appropriate under current conditions. The purpose of this inquiry was (1) to propose the use of subjective probability as a theoretical model for enrollment forecasting, and (2) to report the results of an application of subjective probability to enrollment forecasting at The University of Toledo.

Probability Defined

Three definitions of probability have evolved into the current literature. The first is deductive logic, which establishes the likelihood of some event E by using a symmetric or geometric consideration of the event. The probability of obtaining a "six" on the single roll of a die is an example of this approach. The second definition is called frequency approach. It states that if the event E is repeated N times, N being large, and the event E is observed to occur n of these N times, then $P(E) = n/N$. A coin is

tossed 1000 times, and a head appears 450 times. It is estimated from this information that the probability of a head is 450/1000.

Subjective probability is the third definition. It describes the degree of confidence that an individual, or group of individuals, place in a certain event. An event considered likely to occur is assigned a probability close 1.0. An event unlikely to occur is assigned a probability close to 0. The possible outcome may be described by a discrete or a continuous probability distribution. The subjective approach is broad and flexible, permitting assignment of probabilities to events for which little or no objective data are available.

A Theoretical Enrollment Forecasting Example Using Subjective Probability

Several of the forecasting techniques suggested in the literature are best employed when the university can be divided into homogeneous subpopulations. The university forecast is obtained by totaling estimates from these subpopulations. The cohort-survival and ratio techniques qualify as examples. The reason for the identification of subpopulations is that factors which affect entering freshman maybe different than those related to continuing juniors, for example. Part-time enrollment may be conditioned by variables different than those of full-time. Thus, by focusing on a subpopulations, the appropriate set of factors can be studied and employed for projections rather than studying the set related to the whole population but in varying ways. The crucial requirement is to identify the most reasonable groupings for the particular projection.

Estimates for any subpopulation, considering the normal distribution as descriptive of the probability of certain enrollment outcomes, can be made using subjective probability. The combining of the separate projections for these groups will result in a distribution which is also normally distributed

with a mean (μ) equal to the sum of the subpopulation means and a variance (σ^2) equal to the sum of the individual variances. These formulations are approximately true even if the subpopualtion distributions are not normally distributed.

Consider University A, which may be divided into three homogeneous and exhaustive colleges, (Engineering, Education, and Fine Arts) as the example. Subjective enrollment estimates are made for each of the three colleges for the upcoming academic year. Assume that the distribution of future enrollments for each college follows the normal probability distribution. The mean of a normal distribution can be established by estimating the point on a continuous curve at which the odds are 50-50 that the actual results will be above or below that point. The standard deviation (σ) can be calculated by selecting another point on the normal probability distribution and estimating the probability of exceeding that point. The use of a standard normal equation will provide an estimate of σ.

The Director of Planning for University A estimates the College of Engineering enrollment is as likely to exceed 1000 as to be less than 1000. Further, it is estimated that the probability of enrollment exceeding 950 is 0.95. This is sufficient information to identify the mean and variance of the distribution of future enrollments in the College of Engineering. The mean is 1000, since the normal distribution is symmetrical. The standard deviation can be estimated by substitution in the standard normal equation. The $P(x>950) = 0.95$ is the same as $P(z>1.645)$, found by using a table of z scores. Substituting these values in the standard normal equation, the following results are obtained:

$$z = \frac{x-\mu}{\sigma}$$

$$\sigma = \frac{x-\mu}{z}$$

$$\sigma = \frac{950.-1000}{-1.645}$$

$$\sigma = 30.395 \text{ and } \sigma^2 = 923.86$$

Therefore, the Planning Director has estimated future enrollment in the College of Engineering to be normally distribution with a mean of 1000 and a variance of 923.86.

Knowledge of the variance allows the forecaster to evaluate the probability of various possible outcomes for the entire forecast. The calculation of the forecast variance requires the following steps.

1. The distribution of headcount enrollment in each college is assumed to be normally distributed.
2. The headcounts forecasted are the means, or expected values, for each of the distributions.
3. The probability of enrollment exceeding a value other than the mean is estimated.
4. The variance for each college is calculated using the standard normal equation.
5. The total variance for the forecast of the university is obtained by adding the variances for each college.

The information necessary to determine the variance in the example is as follows:

College	Expected Enrollment	Point on Normal	Prob. of Exceeding Pt.	Estimate of Variance
Engineering	1000	950	0.95	923.86
Education	2000	2040	0.10	976.56
Fine Arts	3000	2940	0.98	856.65
Total	6000			2757.05

Suppose that an administrator is interested in knowing the probability of the headcount enrollment exceeding 6100 students. Using the standard normal equation and the information just determined, this probability can be determined as follows:

$$z = \frac{x - \mu}{\sigma}$$

where

x is the value in question (6100)

μ is the population mean (6000)

σ is the estimated standard deviation

(σ^2 = 2757.05 = 52.51).

The z value is:

$$z = \frac{6100-6000}{52.51} = 1.90$$

The probability of exceeding 1.90 in the standard normal distribution is 0.0287. Thus, the probability of headcount enrollment exceeding 6100 is 2.87%

Conversion to Full-Time Equivalent (FTE) Enrollment

Two universities which have the same headcount enrollment may differ widely in instructional demands. A common expression employed to measure the impact of student course registrations is to determine the full-time equivalent (FTE) enrollment. To obtain the number of FTE's at a university, the number of students in each class is multiplied by the credit hours for which each student is registered. This product is summed over all courses within the University and then divided by 15.

The cost for instruction often depends upon the level of program support required. A graduate course in physics usually is more expensive to teach than a course in introductory psychology. To account for these differences, funding agencies often classify courses according to instructinal needs.

The FTE enrollments are usually classified according to subsidy levels, therefore, forecasts are necessary by subsidy level.

A numerical procedure can be described to convert headcount enrollment by college to an FTE basis by subsidy level. To illustrate this conversion, assume University A consisted of three colleges (Engineering, Education, and Fine Arts) with enrollment for an upcoming period estimated as before by college to be 1,000, 2,000 and 3,000 respectively. Assume further that the state is funding the university at three subsidy levels (undergraduate, master's and doctoral).

The conversion process involves the creation of two matrices based upon a study of the following information.

1. The average number of credit hours taken by a student by college.
2. The proportion of student credit hours in each subsidy level by college.

If the average number of credit hours in the colleges were

College	Average Number of Credit Hours
Engineering	8
Education	9
Fine Arts	8

then multiplying the headcount enrollment in each college by 8/15, 9/15, and 8/15 respectively, the headcount would be converted to FTE by college (533.33, 1200, and 1600).

The proportion of student credit hours in each subsidy category by college is shown below.

	Subsidy Levels		
Colleges	Undergraduate	Master's	Doctoral
Engineering	0.50	0.25	0.25
Education	0.45	0.35	0.20
Fine Arts	0.50	0.40	0.10

Sufficient information has now been provided to complete the conversion process. The FTE at each subsidy level is obtained by multiplying the FTE by college values by the various proportions of student credit hours at the undergraduate levels. The undergraduate total FTE is 1606.7 (533.3·0.50 + 1200·45 + 1600 · 0.50). The master's and doctoral subsidies are obtained in a similar fashion. The FTE values are 1193.3 for the master's level and 533.3 for the doctoral level. The total FTE is 3333.3.

	Headcount		Average Load		Probability Matrix			FTE Matrix		
Engineering	1000	x	8/15	x	0.50	0.25	0.25	266.7	133.3	133.3
Education	2000		9/15		0.45	0.35	0.20	540.0	420.0	240.0
Fine Arts	3000		8/15		0.50	0.40	0.10	800.0	640.0	160.0

The FTE by subsidy level totals are obtained by summing the columns in the FTE matrix.

Variance of the Headcount and FTE Distributions

Knowledge of the variance of the headcounts allow the forecaster to evaluate the probability of various outcomes as well as that of making comparisons about the efficiency of the techniques. The standard deviation of the headcount distribution for University A was previously computed by using the normality assumption to determine the mean and standard deviation for each of the colleges. The variance of the FTE distribution can be determined by performing a linear transformation on the variance of the headcount enrollment by college. The transformation is of the form that if some value "a" is multiplied by each point of the distribution, the variance of the resultant distribution is $a^2 \sigma^2$. The steps in the conversion process are the same as those to convert the headcounts by college to FTE's by subsidy level, except that the squares of the various terms are requried. That is, the squares of

the ratios of the credit hour loads by college to 15 are used ($(8/15)^2$, $(9/15)^2$, and $(8/15)^2$) and the values within the probability matrix are squared. The entire process can be described by matrix algebra in the following steps.

	Headcounts		Square of Avg. Load		Square of Probability Matrix				FTE Variance Matrix		
Engineering	923.86	*	0.2844	*	0.2500	0.0625	0.0625	=	65.69	16.42	16.42
Education	976.56		0.3600		0.2025	0.1225	0.0400		71.19	43.07	14.06
Fine Arts	856.63		0.2844		0.2500	0.1600	0.0100		60.91	38.98	2.44

The terms in the rows and columns of the FTE variance matrix are totaled to determine the variance of the FTE distribution. The total variance is 329.18 and the standard deviation is 18.14. The probability of exceeding 3100 FTE's can now be estimated as follows:

$$z = \frac{x - \mu}{\sigma}$$

where

x is the value in question 3100

μ is the population FTE mean (3333.3)

σ is the estimated standard deviation (18.14)

The z value is:

$$z = \frac{3100-333.3}{18.14} = 1.84$$

The probability of exceeding -1.84 in the standard normal distribution is 0.9671. Thus, the probability of exceeding 3100 FTE is nearly 97 percent.

University of Toledo Trial

The subjective probability approach was employed to project enrollment at The University of Toledo for the fall quarter of 1978. The initial step in developing the forecast was to consult with the Deans of the several colleges to obtain their inputs as to expected headcount enrollments. Next inputs were

obtained from various universities administrators and committees. This information was synthesized into a subjective probability distribution for full-time and part-time, graduate and undergradate headcount enrollment by college. That is, four separate projections were obtained for each of the 8 colleges within The University of Toledo. The means and variances of these 32 separate headcount projections were totaled to determine the mean and variance of headcount distribution for The University of Toledo. As described in the previous section the headcount information was converted to a FTE basis and the FTE at each subsidy level determined. (In Ohio there are 17 subsidy levels, although The University of Toledo does not have enrollment at all levels). The mean and variance of the FTE distribution was determined using the linear transformation on the variance of the headcount distribution.

The following statements briefly summarize the results.

1. The percentage difference between the actual and the forecasted head count enrollment was 0.4 percent. The absolute error was seventy students on a total enrollment of nearly 17,500 students.
2. The percentage difference between actual and forecasted FTE enrollment was 0.3 percent. The absolute error was forty-three FTEs on a total of more than 12,800 FTEs.
3. To support the trial at The University of Toledo most of the conventional forecasting techniques were employed. The forecasting errors using subjective probability were the smallest.

Summary and Extentions to Future Research

Subjective probability can be used as an enrollment forecasting technique for both headcount and FTE using the following steps.

1. Assume the various subpopulation are normally distributed.
2. Determine the mean and standard deviations of these subpopulations.

3. Sum the means and variances of the subpopulations to obtain the mean and variance of the university forecast.
4. Perform a linear transformation on the headcount distribution to obtain the FTE distribution.

The major advantages to this method are:

1. allows the study of homogenious subpopulation
2. allows comparisons with other methods since the variability is known
3. Allows probability statements regarding various headcount and FTE outcomes.

While in the application of subjective probability to enrollment forecasting at The University of Toledo provided reasonable results, the procedure needs to be applied to other institutions in other settings. The University of Toledo is an urban university with enrollment of about 17,500 students and recent years have seen moderate increases in enrollment. Trials where enrollment was not increasing, at a private institution, at a residential college, at a community college, and at a larger university all would provide further insight into the validity of the technique in forecasting. In addition, the method should be repeated several years at The University of Toledo to provide a firmer basis for the subjective estimates.

The assessor's judgments regarding enrollment forecasting are also in need of further study. Of particular interest is: the type of information that is most appropriate to provide the assessor: the question of whether assessors will project declines in enrollment, and the preparation necessary for a person to serve as an assessor.

References

Chernoff, Herman and Moses, Lincoln E. Elementary Decision Theory. New York: John Wiley and Sons, Inc., 1959.

Freund, John E. Mathematical Statistics. 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Harnett, Donald L. Introduction to Statistical Methods. 2nd ed. Reading, MA. Addison-Wesley, 1975.

Lapin, Lawrence. Quantitative Methods for Business Decisions. New York: Harcourt Brace Jovanovich, 1976.

Wing, Paul. Higher Education Enrollment Forecasting: A Manual for State-Level Agencies. Boulder, Colorado: National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, 1974.

Winkler, Robert L. Introduction to Bayesian Inference and Decision. New York: Holt, Rinehart, and Winston, 1972.